

AMENDMENT TO THE CLAIMS

1. (currently amended) A computer-implemented method of extracting information from an information source comprising a plurality of documents, comprising:

accessing strings of text in the information source;

comparing the strings of text in the information source with generalized extraction patterns and identifying a plurality of strings in the information source that match at least one generalized extraction pattern, the generalized extraction patterns including related elements pertaining to a subject, words and wildcards, wherein the at least one word and at least one wildcard are positioned between the related elements and wherein wildcards denote that at least one word and less than a specified number of words in an individual string can be skipped in order to match the individual string to an individual generalized extraction pattern;

extracting a first set of related elements of text pertaining to the subject from a first string of the plurality of strings based on the related elements pertaining to the subject in the at least one generalized extraction pattern, the first string being associated with a first document in the plurality of documents;

extracting a second set of related elements of text pertaining to the subject from a second string of the plurality of strings based on the related elements in the at least one generalized extraction pattern, the second string being associated with a second document in the plurality of documents, wherein at least one of the related elements of text in the first set of related elements is different from each of the related elements of text in the second set of related elements of text; and

outputting the first related set of related elements and the second set of related elements.

2. (Previously Presented) The computer-implemented method of claim 1 and further comprising processing the first related set of elements and the second set of related elements to analyze data in the information source.

3. (previously presented) The computer-implemented method of claim 1 wherein for at least one of the corresponding elements in each of the generalized extraction patterns, there is at least one word positioned between said at least one of the corresponding elements and the wildcards.

4. (cancelled)

5. (currently amended) A computer-readable storage medium for extracting information from an information source comprising a plurality of documents, comprising:

a data structure including a set of generalized extraction patterns including related elements pertaining to a subject, words and an indication of a position for at least one optional word and less than a specified number of words such that the words and the at least one optional word are positioned between the related elements; and

an extraction module using the set of generalized extraction patterns to match a first string and a second string in the information source with one of the generalized extraction patterns, the first string associated with a first document in the plurality of documents and the second string associated with a second document in the plurality of documents, extract a first set of related elements of text pertaining to the subject from the first string based on the related elements in said one of the generalized extraction patterns and a second set of related elements of text pertaining to the subject from the second string based on the related elements in said one of the generalized extraction patterns, wherein at least one of the related elements of text in the first set of related elements is different from each of the related elements of text in the second set of related elements of text, and output the first ~~related~~ set of related elements and the second ~~related~~ set of related elements.

6. (currently amended) The computer-readable storage medium of claim 5 and further comprising a module adapted to process the first set of related elements of text and the second set of related elements of text.

7. (currently amended) The computer-readable storage medium of claim 5 wherein for the generalized extraction patterns there is at least one word positioned between at least one of the elements and the indication.

8 - 24 (cancelled)

25. (previously presented) The computer-implemented method of claim 1 wherein each of the elements of the first set of related elements of text are different from each of the elements of the second set of related elements of text.

26. (previously presented) The computer-implemented method of claim 1 wherein the corresponding related set of elements refer to general elements pertaining to the subject and the first set of related elements and the second set of related elements refer to specific text associated with the general elements.

27. (previously presented) The computer-implemented method of claim 26 wherein the corresponding related set of general elements include at least one of a company/product pair, a book title/author pair, an inventor/invention information pair and a question/answer pair.

28. (previously presented) The computer-implemented method of claim 27 wherein the first set of related elements and the second set of related elements refer to at least one of a specific company, a specific product, a specific book title, a specific author, a specific inventor, a specific invention, a specific question and a specific answer.

29. (previously presented) The computer-implemented method of claim 1 wherein the plurality of documents include at least one of a collection of documents, news articles and a collection of customer feedback.

30. (currently amended) The computer-readable storage medium of claim 5 wherein each of the elements of the first set of related elements of text are different from each of the elements of the second set of related elements of text.

31. (currently amended) The computer-readable storage medium of claim 5 wherein the corresponding related set of elements refer to general elements pertaining to the subject and the first set of related elements and the second set of related elements refer to specific text associated with the general elements.

32. (currently amended) The computer-readable storage medium of claim 31 wherein the corresponding related set of general elements include at least one of a company/product pair, a book title/author pair, an inventor/invention information pair and a question/answer pair.

33. (currently amended) The computer-readable storage medium of claim 32 wherein the first set of related elements and the second set of related elements refer to at least one of a specific company, a specific product, a specific book title, a specific author, a specific inventor, a specific invention, a specific question and a specific answer.

34. (currently amended) The computer-readable storage medium of claim 5 wherein the plurality of documents include at least one of a collection of documents, news articles and a collection of customer feedback.